一种改进的局部线性回归估计器及其在雷达高度计海况偏差估计中的应用

蒋茂飞[①②③]　许　可*[①②]　刘亚龙[④]　王磊[①②]

[①](中国科学院微波遥感技术重点实验室　北京　100190)

[②](中国科学院国家空间科学中心　北京　100190)

[③](中国科学院大学　北京　100049)

[④](国家海洋局烟台海洋环境监测中心站　烟台　264000)

摘　要：在建立雷达高度计海况偏差(Sea State Bias, SSB)非参数模型时，通常会用到局部线性回归(Local Linear Regression, LLR)估计器，而传统的局部线性回归估计器涉及高维矩阵运算，当建模的数据量较大时，估计海况偏差需要大量的时间，从而使得非参数估计方法很难用于高维海况偏差模型。该文提出一种改进的局部线性回归(Improved Local Linear Regression, ILLR)估计器，可以避免传统的LLR估计器所需的高维矩阵运算，在不影响海况偏差估计结果的条件下，将局部线性回归估计器获取加权函数的时间复杂度由$O(N^2)$降低为$O(N)$，从而大幅地降低估计海况偏差所需的时间，为实现高维非参数海况偏差模型的实时运算奠定了基础。

关键词：雷达高度计；海况偏差；非参数估计；LLR估计器

中图分类号：TN953　　**文献标识码**：A　　**文章编号**：

DOI：10.11999/JEIT151280

Improved Local Linear Regression Estimator and Its Application to the Estimation for Radar Altimeter Sea State Bias

JIANG Maofei[①②③]　XU Ke[①②]　LIU Yalong[④]　WANG Lei[①②]

[①](*Key Laboratory of Microwave Remote Sensing, Chinese Academy of Sciences, Beijing 100190, China*)

[②](*National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China*)

[③](*University of Chinese Academy of Sciences, Beijing 100049, China*)

[④](*Yantai Marine Environmental Monitoring Center Station, State Oceanic Administration, Yantai 264000, China*)

Abstract: The Local Linear Regression (LLR) estimator is usually used when developing a nonparametric model for radar altimeter Sea State Bias (SSB). However, the conventional LLR estimator contains matrices with high dimension. When a large number of data are used to develop the SSB model, the SSB estimation costs too much time. Therefore, the nonparametric estimation method can hardly be used to develop high-dimensional SSB models. This paper presents an improved LLR estimator, which can avoid high-dimensional matrix operations. The improved LLR estimator can greatly reduce the time for SSB estimation without affecting the estimated accuracy. So the improved LLR estimator can laid the foundation for high-dimensional SSB models.

Key words: Radar altimeter; Sea state bias; Nonparametric estimation; LLR estimator

1 引言

全球海平面上升已越来越引起人们的关注，雷达高度计的一个重要应用就是测量平均海面高度[1, 2]。但是由于海况偏差的存在，高度计测量得到的平均海面会低于实际的平均海面，必须对它进行校正。海况偏差主要是由于海面的非高斯分布特性引起，包括电磁偏差和偏斜度偏差[3]。利用现有方法得到的海况偏差的不确定度能够达到2 cm[4]，已经成为Jason系列测高卫星最大的误差源[5, 6]。

目前，业务化运行的雷达高度计的海况偏差校正主要采用以高度计测得的风速和有效波高作为输入的非参数SSB模型[6,7]。非参数SSB模型的构建通常需要用到局部线性回归估计器，而传统的局部线性回归估计器需要大量的矩阵运算，从而消耗大量的运算时间，极大地限制了非参数估计在高维SSB模型中的应用。

本文提出了一种改进的局部线性回归估计器，该估计器避免了高维矩阵运算，将从局部线性回归估计器获取加权函数的时间复杂度由$O(N^2)$降低为$O(N)$。数值实验结果表明，在其它条件都相同的条件下，和传统的局部线性回归估计器相比，改进的局部线性回归估计器不会影响SSB估计的结果，但可以大幅地减低估计SSB所需的时间。

收稿日期：2015-11-17

*通信作者：许可　xuke@mirslab.cn

2 局部线性回归估计器及其在海况偏差非参数估计中的应用

GASPAR 等人[8]在 1998 年首次将非参数估计的方法用于 SSB 模型的构建，和参数模型[4, 9]相比，非参数模型的精度更高，从而被更广泛地应用于高度计测量海面高度的海况偏差校正。用于非参数 SSB 模型的估计器主要有 Nadaraya-Watson (NW)估计器[8]和局部线性回归估计器[10]，NW 估计器在接近边界处会出现较大偏差，而局部线性回归估计器则可以很好解决这一问题[11,12]，所以局部线性回归估计器得到了更广泛的使用。

2.1 局部线性回归估计器

给定 N 组观测值$(y_i,\boldsymbol{x}_i)$，$\boldsymbol{x}_i$ 是含有 p 个变量的向量，$\boldsymbol{x}_i$ 和y_i 满足：

$$y_i = r(\boldsymbol{x}_i) + e_i \tag{1}$$

其中r 是回归函数，e_i 是一个零均值的误差项。对于回归函数$r(\boldsymbol{x})$，在$\boldsymbol{x}$ 的邻域$\boldsymbol{z}$ 对其进行 Taylor 展开有：

$$r(\boldsymbol{z}) = b_0 + \sum_{j=1}^{p} b_j(z_j - x_j) \tag{2}$$

对于给定的 N 组观测值$(y_i,\boldsymbol{x}_i)$，可以通过加权最小二乘的方法找到一组$\boldsymbol{b} = (b_0,b_1,\ldots,b_p)^{\mathrm{T}}$，使得式(3)最小：

$$\sum_{i=1}^{N}\Big[y_i - b_0 - \sum_{j=1}^{p} b_j(x_{ij} - x_j)\Big]^2 K_H(\boldsymbol{x}_i - \boldsymbol{x}) \tag{3}$$

式中，$K_H(\boldsymbol{x}_i - \boldsymbol{x})$是加权函数，对于局部线性回归估计器，$K_H(\boldsymbol{x}_i - \boldsymbol{x})$通常选用核函数，常用的核函数有高斯核函数和 Epanechnikov 核函数。令 $\boldsymbol{X}_D = \begin{bmatrix} 1 & x_{11}-x_1 & \cdots & x_{1p}-x_p \\ 1 & x_{21}-x_1 & \cdots & x_{2p}-x_p \\ \vdots & \vdots & & \vdots \\ 1 & x_{N1}-x_1 & \cdots & x_{Np}-x_p \end{bmatrix}$，$\boldsymbol{y}^{\mathrm{T}} = [y_1,y_2,\ldots,y_N]$，$\boldsymbol{W} = \mathrm{diag}\{K_H(\boldsymbol{x}_i - \boldsymbol{x})\}$是一个$(N \times N)$的加权矩阵，式(3)可以改写为：

$$(\boldsymbol{y} - \boldsymbol{X}_D\boldsymbol{b})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{X}_D\boldsymbol{b}) \tag{4}$$

最小化式(4)，可以得到：

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D)^{-1}\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y} \tag{5}$$

$r(\boldsymbol{x})$的局部线性回归估计器$\hat{r}_{\mathrm{LLR}}(\boldsymbol{x})$为$\hat{b}_0$，而$\hat{\boldsymbol{b}}$的其它成分是函数$r$ 的一阶偏导数，即

$$\hat{r}_{\mathrm{LLR}}(\boldsymbol{x}) = \hat{b}_0 = \boldsymbol{e}_1^{\mathrm{T}}(\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D)^{-1}\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y} \tag{6}$$

式中,$\boldsymbol{e}_1$是一个含有 p+1 个变量的单位向量，$\boldsymbol{e}_1^{\mathrm{T}} = [1,0,\ldots,0]$。

对于给定的 N 组观测值$(y_i,\boldsymbol{x}_i)$，根据非参数估计理论，条件期望$E[y\,|\,\boldsymbol{x}]$表示在给定$\boldsymbol{x}$ 的条件下y 的均值，其核估计器$\hat{r}(\boldsymbol{x})$可以看作是观测值y_i 的加权平均：

$$\hat{r}(\boldsymbol{x}) = \sum_{i=1}^{N} y_i a_N(\boldsymbol{x} - \boldsymbol{x}_i) \tag{7}$$

对于局部线性回归估计器$\hat{r}_{\mathrm{LLR}}(\boldsymbol{x})$，在式(6)中，令$\boldsymbol{T} = \boldsymbol{e}_1^{\mathrm{T}}(\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D)^{-1}\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}$，就可以得到权函数：

$$a_N(\boldsymbol{x} - \boldsymbol{x}_i) = \boldsymbol{T}(i) \tag{8}$$

2.2 局部线性回归估计器在海况偏差非参数估计中的应用

通常使用交叉点处的海面高度不符值来建立SSB模型，卫星绕地球运行一周的轨迹包括两个弧段(Pass)，分别为上升轨道和下降轨道，上升轨道和下降轨道的交点就称为交叉点，而交叉点海面高度不符值是指上升轨道和下降轨道在交叉点处测得的海面高度之差[13]。

SSH' 表示未经 SSB 校正的海面高度测量值，假设 SSB 是输入变量$\boldsymbol{x} = (U,\mathrm{SWH})$的函数$j$，并令$y = \mathrm{SSH}_2' - \mathrm{SSH}_1'$，根据文献[8]和文献[10]可知：

$$y = j(\boldsymbol{x}_2) - j(\boldsymbol{x}_1) + e \tag{9}$$

式中，e 是一个零均值的误差项，角标 1 和角标 2 分别对应上升轨道和下降轨道。在给定$\boldsymbol{x}_2 = \boldsymbol{x}$ 的条件下，

y 的条件期望为：

$$E[y\,|\,\boldsymbol{x}_2=\boldsymbol{x}]=\varphi(\boldsymbol{x})-E[\varphi(\boldsymbol{x}_1)\,|\,\boldsymbol{x}_2=\boldsymbol{x}] \tag{10}$$

对于 N 组测量值$(y_i,\boldsymbol{x}_{1i},\boldsymbol{x}_{2i})$，利用式（7）和式（10）可以得到：

$$\varphi(\boldsymbol{x})=\sum_{i=1}^{N}y_i a_N(\boldsymbol{x}-\boldsymbol{x}_{2i})+\sum_{i=1}^{N}\varphi(\boldsymbol{x}_{1i})a_N(\boldsymbol{x}-\boldsymbol{x}_{2i}) \tag{11}$$

式(11)给出了估计$\varphi(\boldsymbol{x})$的方法，只要知道了$\varphi(\boldsymbol{x}_{1i})$，就可以估计出任意输入$\boldsymbol{x}$ 所对应的 SSB 估计值$\varphi(\boldsymbol{x})$。利用$\boldsymbol{x}_{1j}$ 替换式(11)中$\boldsymbol{x}$，式(11)可以重写为：

$$\varphi(\boldsymbol{x}_{1j})=\sum_{i=1}^{N}y_i a_N(\boldsymbol{x}_{1j}-\boldsymbol{x}_{2i})+\sum_{i=1}^{N}\varphi(\boldsymbol{x}_{1i})a_N(\boldsymbol{x}_{1j}-\boldsymbol{x}_{2i}),\ \forall j=1,\ldots,N \tag{12}$$

将式(12)转化为矩阵的形式可以得到：

$$(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\varphi}_1=\boldsymbol{A}\boldsymbol{y} \tag{13}$$

式中，$\boldsymbol{I}$ 和$\boldsymbol{A}$ 都是$(N\times N)$的矩阵，矩阵$\boldsymbol{I}$ 是一个单位矩阵，而矩阵$\boldsymbol{A}$ 的元素为$a_{ji}=a_N(\boldsymbol{x}_{1j}-\boldsymbol{x}_{2i})$，$\boldsymbol{\varphi}_1^{\mathrm{T}}=[\varphi(\boldsymbol{x}_{11}),\varphi(\boldsymbol{x}_{12}),\ldots,\varphi(\boldsymbol{x}_{1N})]$，$\boldsymbol{y}^{\mathrm{T}}=[y_1,y_2,\ldots,y_N]$。

矩阵$\boldsymbol{I}-\boldsymbol{A}$ 是一个奇异矩阵，其秩为 N-1。为了消除不确定性，可以先给$\boldsymbol{\varphi}_1$ 中的第一个元素一个假设值：$\varphi(\boldsymbol{x}_{11})=\varphi_0$，那么式(13)可以重写为：

$$\boldsymbol{B}_1\boldsymbol{\varphi}=\boldsymbol{A}\boldsymbol{y}-\boldsymbol{B}_0\varphi_0 \tag{14}$$

式中，$\boldsymbol{\varphi}$ 包含$\boldsymbol{\varphi}_1$ 除$\varphi(\boldsymbol{x}_{11})$之外的 N-1 个元素，即$\boldsymbol{\varphi}^{\mathrm{T}}=[\varphi(\boldsymbol{x}_{12}),\ldots,\varphi(\boldsymbol{x}_{1N})]$，$\boldsymbol{B}_0$是矩阵$\boldsymbol{I}-\boldsymbol{A}$ 的第 1 列，$\boldsymbol{B}_1$是其余各列组成的矩阵，利用最小二乘法求解$\boldsymbol{\varphi}$：

$$\hat{\boldsymbol{\varphi}}=(\boldsymbol{B}_1^{\mathrm{T}}\boldsymbol{B}_1)^{-1}\boldsymbol{B}_1^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{B}_0\varphi_0) \tag{15}$$

再联合$\varphi(\boldsymbol{x}_{11})$，就可以得到$\boldsymbol{\varphi}_1$，将它带入式(11)就可以得到任意$\boldsymbol{x}$ 所对应的$\varphi(\boldsymbol{x})$的估计值。

在估计 SSB 时，可制作一个 SSB 估计值关于风速 U、有效波高 SWH 的 2 维查找表，再通过插值的方法得到任意高度计测量点的 SSB 估计值。为了得到 SSB 估计值的查找表，通常是先利用一个 cycle 的数据计算查找表中每个网格点的输入变量$\boldsymbol{x}$ 所对应的 SSB 估计值$\varphi(\boldsymbol{x})$，再对多个 cycle 得到的结果进行平均。每个 cycle 的交叉点的有效个数一般可达 5000~7000 个，如果将所有交叉点都用于建模，那么式(15)中矩阵求逆需要大量的运算，消耗大量计算时间。在文献[8]的模型中，从每个 cycle 随机抽取 500 个有效的交叉点来建立 SSB 估计模型。为了避免了式(15)中矩阵求逆运算，GASPAR 等人[10]在 2002 年利用 LSQR 算法来求线性方程式(14)的解，并将每个 cycle 的所有交叉点数据都用于 SSB 的建模。LSQR 算法是 PAIGE 等人[14]在 1982 年提出的一种解稀疏线性方程组的迭代方法，它被用于求线性方程组$Ax=b$的解或者使$\|Ax-b\|_2$ 最小的解。为了进一步提高求解线性方程式(14)的速度，本文使用了 LSMR 算法来解方程式(14)，LSMR 算法是 FONG 等人[15]在 2010 年提出的一种解线性方程组的迭代算法。和 LSQR 算法一样，LSMR 算法也是用于求线性方程组$Ax=b$的解或者使$\|Ax-b\|_2$ 最小的解，但和 LSQR 算法相比，使用 LSMR 算法可以更快和更安全的达到收敛条件。

2.3 改进的局部线性回归估计器

非参数 SSB 模型最终得到的是一个关于风速 U 和有效波高 SWH 的 2 维查找表，需要计算查找表中每个网格点的输入$\boldsymbol{x}=(U,\mathrm{SWH})$所对应的 SSB 估计值。虽然利用 LSQR 和 LSMR 算法来解线性方程式(14)可以避免式(15)式矩阵求逆运算，但式(11)中的加权函数$a_N(\boldsymbol{x}-\boldsymbol{x}_{2i})$和式(12)中加权矩阵 $\mathbf{A}$ 中的元素$a_N(\boldsymbol{x}_{1j}-\boldsymbol{x}_{2i})$都是通过局部线性回归估计器得到，从式(5)和式(8)也可以看出通过局部线性回归估计器得到加权函数时也需要用到高维矩阵运算，本文希望避免高维矩阵运算，在对结果影响不大的条件下，能大大降低运算时间。

对于传统的利用风速(U)和有效波高(SWH)来建立 SSB 估计模型时，$\boldsymbol{x}=(x_1,x_2)=(U,\mathrm{SWH})$，$\boldsymbol{z}$ 在$\boldsymbol{x}$ 的邻域，回归函数r 的局部线性模型可以表示为：

$$r(\boldsymbol{z})=b_0+b_1(z_1-x_1)+b_2(z_2-x_2) \tag{16}$$

通过找到 $\boldsymbol{b} = (b_0, b_1, b_2)^{\mathrm{T}}$，使得式(17)的函数最小：

$$Z = \sum_{i=1}^{N} \left[y_i - b_0 - b_1(x_{i1} - x_1) - b_2(x_{i2} - x_2) \right]^2 K_H(\boldsymbol{x}_i - \boldsymbol{x}) \tag{17}$$

令 $\frac{\partial Z}{\partial b_0} = 0$，$\frac{\partial Z}{\partial b_1} = 0$，$\frac{\partial Z}{\partial b_2} = 0$，那么可以得到一个方程组：

$$\left.\begin{aligned} S_{00}b_0 + S_{10}b_1 + S_{01}b_2 &= a \\ S_{10}b_0 + S_{20}b_1 + S_{11}b_2 &= b \\ S_{01}b_0 + S_{11}b_1 + S_{02}b_2 &= c \end{aligned}\right\} \tag{18}$$

其中，a, b, c 和 S_{mn} 可以表示为：

$$\left.\begin{aligned} a &= \sum_{i=1}^{N} y_i K_H(\boldsymbol{x} - \boldsymbol{x}_i) \\ b &= \sum_{i=1}^{N} y_i (x_{i1} - x_1) K_H(\boldsymbol{x} - \boldsymbol{x}_i) \\ c &= \sum_{i=1}^{N} y_i (x_{i2} - x_2) K_H(\boldsymbol{x} - \boldsymbol{x}_i) \\ S_{mn} &= \sum_{i=1}^{N} (x_{i1} - x_1)^m (x_{i2} - x_2)^n K_H(\boldsymbol{x} - \boldsymbol{x}_i) \end{aligned}\right\} \tag{19}$$

令 $\boldsymbol{S} = \begin{bmatrix} S_{00} & S_{10} & S_{01} \\ S_{10} & S_{20} & S_{11} \\ S_{01} & S_{11} & S_{02} \end{bmatrix}$，方程组式(18)可以表示为 $\boldsymbol{S}\boldsymbol{b} = (a, b, c)^{\mathrm{T}}$，这里用伪逆来求 $\boldsymbol{b}$：

$$\hat{\boldsymbol{b}} = \mathrm{pinv}(\boldsymbol{S}) \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{20}$$

$$\hat{b}_0 = \boldsymbol{e}_1^{\mathrm{T}} \cdot \mathrm{pinv}(\boldsymbol{S}) \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{21}$$

令 $\boldsymbol{Q} = \boldsymbol{e}_1^{\mathrm{T}} \cdot \mathrm{pinv}(\boldsymbol{S})$，那么

$$\begin{aligned} \hat{b}_0 &= \boldsymbol{Q}(1) \cdot a + \boldsymbol{Q}(2) \cdot b + \boldsymbol{Q}(3) \cdot c \\ &= \sum_{i=1}^{N} \left[\boldsymbol{Q}(1) + \boldsymbol{Q}(2) \cdot (x_{i1} - x_1) + \boldsymbol{Q}(3) \cdot (x_{i2} - x_2) \right] K_H(\boldsymbol{x}_i - \boldsymbol{x}) \cdot y_i \end{aligned} \tag{22}$$

所以，权值函数可以表示为：

$$a_N(\boldsymbol{x}_i - \boldsymbol{x}) = \left[\boldsymbol{Q}(1) + \boldsymbol{Q}(2) \cdot (x_{i1} - x_1) + \boldsymbol{Q}(3) \cdot (x_{i2} - x_2) \right] K_H(\boldsymbol{x}_i - \boldsymbol{x}) \tag{23}$$

将式(23)替换到式(8)，就可以将 ILLR 估计器用于非参数 SSB 模型的构建，和 LLR 估计器相比，ILLR 估计器避免了高维矩阵运算。

由式(8)和式(23)可知，两种估计器得到的权函数有所区别，ILLR 估计器和 LLR 估计器在本质上是相同的，都是通过对式(3)求偏导，并使其偏导为零得到，如果将式(5)重写为 $\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D\boldsymbol{b} = \boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y}$，其展开后就得到了式(18)，所以只要 $\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D$ 不是奇异矩阵，ILLR 估计器和 LLR 估计器得到的结果就完全相同。在 LLR 估计器中，由于 $\boldsymbol{W}$ 是对角矩阵，$\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}$ 和 $\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}$ 在运算过程中会产生很多结果为零值,它们会继续参与 $\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D$ 和 $\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y}$ 的矩阵运算，而在 ILLR 估计器中，LLR 估计器中 $\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D$ 和 $\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y}$ 在计算过程中产生的那些中间结果为零的值不会参与式(18)，式(9)的运算，这就大大节约了计算时间。从算法时间复杂度方面分析，在 LLR 估计器中，权函数 $a_N(\boldsymbol{x}_i - \boldsymbol{x})$ 由 $\boldsymbol{T} = \boldsymbol{e}_1^{\mathrm{T}}(\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D)^{-1}\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}$ 得到，$\boldsymbol{e}_1^{\mathrm{T}}$ 是 (1×3) 的矩阵，$\boldsymbol{X}_D$ 是

$(N\times 3)$ 的矩阵，$\boldsymbol{W}$ 是 $(N\times N)$ 的矩阵，$\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}$ 是 $(3\times N)$ 的矩阵，$\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D$ 是 (3×3) 的矩阵，所以 $\boldsymbol{X}_D^{\mathrm{T}}$ 和 $\boldsymbol{W}$ 相乘需要 $3N^2$ 次乘法运算，$\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}$ 和 $\boldsymbol{X}_D$ 相乘需要 3^2N 次乘法运算，矩阵 $\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D$ 求逆需要 3^3 次乘法运算，$(\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D)^{-1}$ 和 $\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}$ 相乘需要 3^2N 次乘法运算，$\boldsymbol{e}_1^{\mathrm{T}}$ 和 $(\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}_D)^{-1}\boldsymbol{X}_D^{\mathrm{T}}\boldsymbol{W}$ 相乘需要 $3N$ 次乘法运算，所以利用 LLR 估计器获得权函数的时间复杂度为 $\mathrm{O}(N^2)$。由式(18)可知，计算 a，b，c 和 S_{mn} 分别需要进行 N，2N，2N 和 $(m+n+1)?N$ 次乘法运算，$\boldsymbol{S}$ 是 (3×3) 的矩阵，所以利用 ILLR 估计器获得权函数的时间复杂度为 $\mathrm{O}(N)$。

3 数值实验

3.1 数据来源和处理

本文使用 Jason-2 高度计的地球物理数据集(GDR)产品，Jason-2 卫星于 2008 年 6 月发射，它是 TOPEX/POSEIDON、Jason-1 测高卫星的后继星，被公认为目前精度最高的测高卫星。本文在建立非参数 SSB 模型时使用了 cycle 51~100 共 50 个 cycle 的数据，而在评估 ILLR 估计器对 SSB 估计结果的影响时使用了 2009(cycle19~cycle55)，2010(cycle56~cycle92)和 2011(cycle93~cycle128)共 3 年的数据。

建立 SSB 模型需要用到风速、有效波高、交叉点处未经 SSB 校正的海面高度不符值。计算海面高度所需的各种校正项都需要插值到交叉点处，本文采用 3 次样条插值的方法，为了得到更好的插值效果，只有在交叉点两侧各有 4 个连续的测量点时得到的交叉点才有效，这样可以保证较好的高度计数据质量。

3.2 核函数和带宽的选择

核函数和带宽的选择对于非参数 SSB 模型非常重要。目前，最常用的核函数有高斯核函数和 Epanechnikov 核函数，而用于非参数 SSB 模型的带宽包括全局带宽和局部带宽两种，全局带宽是指对于所有的输入 $\boldsymbol{x}=(U,\mathrm{SWH})$，风速的带宽 h_U 和有效波高的带宽 h_{SWH} 都为固定值，而局部带宽是指 h_U 和 h_{SWH} 会根据数据密度自动进行调节。

和高斯核函数相比，Epanechnikov 核函数可以将线性系统式(14)中矩阵 $\boldsymbol{B}_1$ 转换稀疏矩阵，从而大大提高了计算效率，但在数据稀疏区域估计的效果比较差。和全局带宽相比，局部带宽可以根据数据的密集程度调节带宽，从而在数据稀疏区域优势较为明显。本文的数值实验只是为了评估改进的局部线性估计器的计算效率以及对估计结果的影响，所以对核函数和带宽的选择没有太多的要求。本文采用了两种核函数和带宽的组合，第 1 种是采用高斯核函数和全局带宽，对应于文献[10]的模型，第 2 种是 Epanechnikov 核函数和局部带宽，对应于文献[12]的模型。

3.3 实验结果

3.3.1 LLR 估计器和 ILLR 估计器估计 SSB 效率比较

首先比较利用 LLR 估计器和 ILLR 估计器估计 SSB 所用的时间，表 1 给出了利用高斯函数和全局带宽得到的结果，随着每次用于 SSB 估计的交叉点数 N 的增大，LLR 估计器里的矩阵的维度就相应地增大，所需的计算时间也会明显增大。ILLR 估计器中矩阵的维度不会随着 N 的增大而改变，虽然其所需的时间也会随着 N 的增大而增大，但其增大的幅度相对较小。表 2 给出了利用 Epanechnikov 核函数和局部带宽估计 SSB 所用的时间，正常情况下，Epanechnikov 核函数可以将线性系统式(14)中的矩阵 $\mathbf{B}_1$ 转换为稀疏矩阵，从而提高计算效率，降低估计 SSB 所需的时间，但由于采用了局部带宽，在估计 SSB 时，对于每一输入 $\boldsymbol{x}$，都需要先确定其所在网格所对应的带宽，这样就会增加计算量，从而使得计算时间增大。和表 1 一样，随着 N 的增大，利用 LLR 估计器估计 SSB 所需的计算时间大幅增加，而利用 ILLR 估计器估计 SSB 所需的计算时间增幅相对较小。从表 1 和表 2 可以看出，和 LLR 估计器相比，利用 ILLR 估计器大可以大幅度地降低估计 SSB 所需的时间，N 越大，改进的效果越明显。

非参数 SSB 模型最终得到的是一个关于 U 和 SWH 的查找表，需要计算每个查找表中每个网格点所对应的 SSB 估计值，通过 U 和 SWH 只能得到有限的 SSB 的信息，将新的参量(如波周期参数)引入 SSB 模型的构建已经成为 SSB 估计的一个新的研究方向[6, 16, 17]。在建立高维非参数 SSB 模型时，最终得到的结果将是一个高维查找表，其所对应的网格点数将远远多于 2 维查找表的网格点数，使用 LLR 估计器所需消耗的时间将非常巨大，并且随着维度的增加，必须要有更多的数据参与模型的构建才能保证模型的有效性，这对于 LLR 估计器是巨大的挑战，而 ILLR 估计器则可以很好地解决这一问题。

表 1 利用高斯核函数和全局带宽估计 SSB 所用的时间

估计器	N=500	N=2000	N=5000
LLR	174 s	37 min	6.8 h
ILLR	28 s	1.5 min	6 min

表 2 利用 Epanechnikov 核函数和局部带宽估计 SSB 所用的时间

估计器	N=500	N=2000	N=5000
LLR	330 s	77 min	12.4 h
ILLR	38 s	2 min	8.9 min

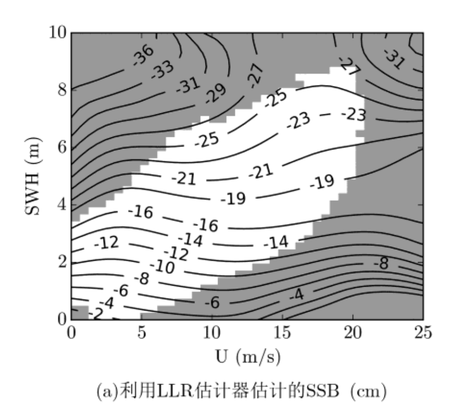
3.3.2 LLR 估计器和 ILLR 估计器估计 SSB 结果比较

图 1 是利用高斯核函数和全局带宽得到的结果，根据数据的密集程度将数据划分为两部分，图中阴影部分表示高度计测量值个数小于 100 的区域，代表数据稀疏区域，而另外一部分则代表数据密集区域。图 1(a)和图 1(b)的等值线分别表示利用 LLR 估计器和 ILLR 估计器得到的 SSB 查找表在(U,SWH)平面的分布，等值线的单位为 cm。很明显，利用两种估计器得到的 SSB 随 U 和 SWH 的变化趋势完全相同，在 U 相同的条件下，SSB 的幅值都随着 SWH 的增大而增大；而在 SWH 相同的条件下，随着 U 的增大，SSB 的幅值都是先增大，而当 U 增大到一定值时，SSB 的幅值都逐渐减小。图 1(c)的等值线表示利用 LLR 估计器和 ILLR 估计器得到的 SSB 之差在(U,SWH)平面的分布，可以看出，在数据密集区域，两种估计器得到的 SSB 之差大约在 10^{-4} cm 量级，基本上可以忽略。

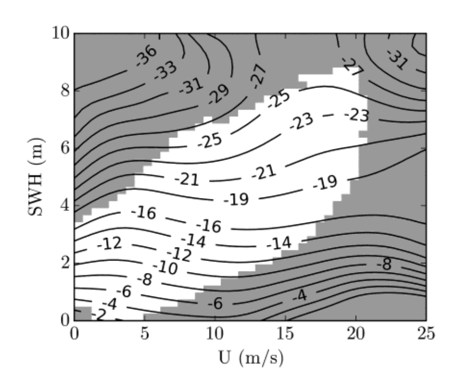
图 2 是利用 Epanechnikov 核函数和局部带宽得到的结果，图 2(a)和图 2(b)的等值线分别表示利用 LLR 估计器和 ILLR 估计器得到的 SSB 查找表在(U,SWH)平面的分布，等值线的单位为 cm，同图 1 一样，利用两种估计器得到的 SSB 随 U 和 SWH 的变化趋势也是相同的，由于使用了 Epanechnikov 核函数，在低风速高海况和高风速低海况的数据稀疏区域，由于使用到的数据比较少，得到的结果不如图 1 中利用高斯核函数得到的结果平滑。图 2(c)的等值线表示利用 LLR 估计器和 ILLR 估计器得到的 SSB 之差在(U,SWH)平面的分布，在数据密集区域，利用两种估计器得到的 SSB 之差大约在 10^{-6} cm 量级，可以认为利用两种估计器得到的 SSB 基本相同。

图 3 给出了利用 LLR 估计器和 ILLR 估计器得到的 SSB 估计值的散点图，图中的结果是通过将两种估计器得到的 SSB 查找表分别用于 2011 年 Jason-2 高度计 SSB 的估计得到，可以看出，两种估计器得到的 SSB 估计值一致性非常好。再将两种估计器得到的 SSB 的查找表分别用于 2009~2011 年 Jason-2 高度计 SSB 的估计，分别利用每年的数据进行统计，表 3 给出了统计结果，当利用高斯核函数和全局带宽时，两种估计器得到的 SSB 估计值之差的最大值在 10^{-3} cm 量级，平均值在 10^{-4} cm 量级；而当利用 Epanechnikov 核函数和局部带宽时，两种估计器得到的 SSB 估计值之差的最大值在 10^{-5} cm 量级，平均值在 10^{-6} cm 量级。所以不管是哪种核函数和带宽的组合，ILLR 估计器对 SSB 的估计结果的影响都可以忽略。

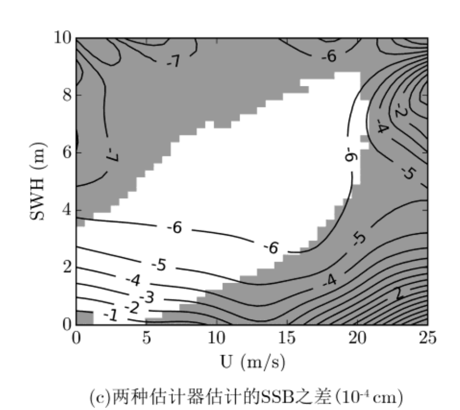
目前，国际上主要采用解释方差来对 SSB 模型进行评估。解释方差是指未经 SSB 校正的海面高度不符值的方差减去经过 SSB 校正后的海面高度不符值的方差，解释方差可以理解为海面高度不符值的方差中 SSB 能够解释的部分，解释方差越大，说明 SSB 模型越有效。对两种不同的核函数和带宽组合，分别使用 2009~2011 3 年的数据计算利用两种估计器得到的 SSB 的解释方差，表 4 给出了利用两种估计器得到的 SSB 的解释方差之差。当使用高斯核函数和全局带宽时，两种估计器得到的 SSB 的解释方差之差在 10^{-5} cm^2 量级，而当使用 Epanechnikov 核函数和局部带宽时，两种估计器得到的 SSB 的解释方差之差在 10^{-9}~10^{-7} cm^2 量级，所以不管是哪种核函数和带宽的组合，和 LLR 估计器相比，ILLR 估计器对 SSB 模型有效性的影响都可以忽略，从而也不会影响高度计最终的测高精度。



(a)利用LLR估计器估计的SSB (cm)

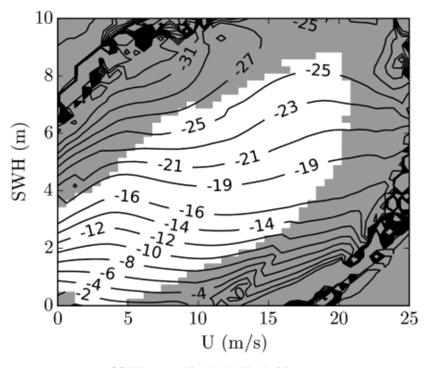


(b)利用ILLR估计器估计的SSB (cm)

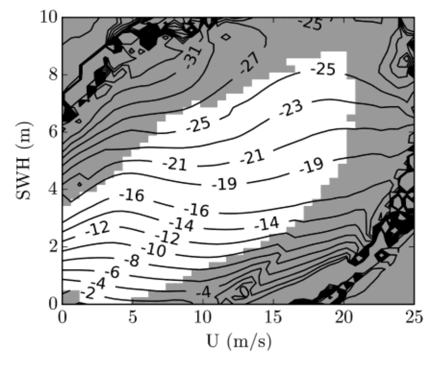


(c)两种估计器估计的SSB之差(10^{-4} cm)

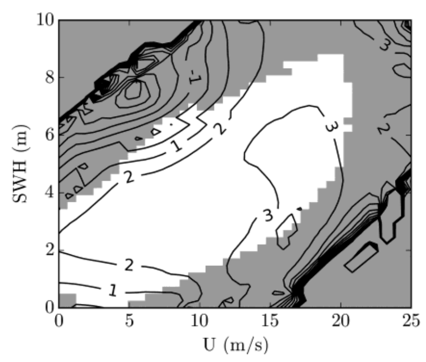
图 1 利用高斯核函数和全局带宽得到的结果在(U,SWH)平面的分布



(a)利用LLR估计器估计的SSB (cm)

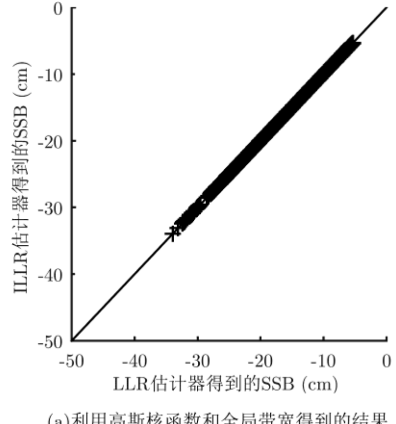


(b)利用ILLR估计器估计的SSB (cm)

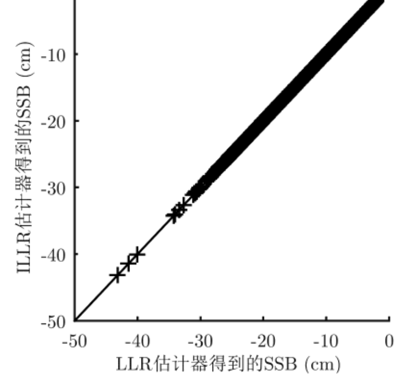


(c)两种估计器估计的SSB之差(10^{-6} cm)

图 2 利用 Epanechnikov 核函数和局部带宽得到的结果在(U,SWH)平面的分布



(a)利用高斯核函数和全局带宽得到的结果



(b)利用Epanechnikov核函数和局部带宽得到的结果

图 3 利用 LLR 估计器和 ILLR 估计器得到的 SSB 估计值的散点图

表 3 利用两种估计器得到的 SSB 之差(cm)统计

核函数和带宽组合	最大值	平均值

	2009 年	2010 年	2011 年	2009 年	2010 年	2011 年
高斯核+全局带宽	1.26×10^{-3}	1.01×10^{-3}	1.13×10^{-3}	5.45×10^{-4}	5.51×10^{-4}	5.49×10^{-4}
Epanechnikov 核+局部带宽	3.63×10^{-5}	8.13×10^{-5}	5.01×10^{-5}	2.39×10^{-6}	2.41×10^{-6}	2.40×10^{-6}

表 4 利用两种估计器得到的 SSB 的解释方差之差(cm^2)

核函数和带宽组合	检验数据集		
	2009 年	2010 年	2011 年
高斯核+全局带宽	4.83×10^{-5}	4.88×10^{-5}	5.27×10^{-5}
Epanechnikov 核+局部带宽	7.88×10^{-8}	1.28×10^{-7}	1.77×10^{-9}

4 结论

非参数模型由于其较高的精度已被广泛用于雷达高度计的 SSB 校正，在利用非参数估计方法估计 SSB 时，通常会采用 LLR 估计器来求加权矩阵里的权函数，而 LLR 估计器涉及高维矩阵运算，当数据量较大时需要大量的计算时间。

本文提出了一种 ILLR 估计器，该估计器可以避免高维矩阵运算，从而极大地提高计算效率。本文分别使用高斯核函数加全局带宽以及 Epanechnikov 核函数加局部带宽这两种组合来估计 SSB，在两种组合中，利用 ILLR 估计器得到的 SSB 和利用 LLR 估计器得到的 SSB 之差基本上可以忽略；ILLR 估计器可以大幅度地降低估计 SSB 所需要的时间，当数据量很大时，这种改进的效果更为明显。目前业务化运行的高度计都是采用基于风速和有效波高的 2 维 SSB 模型，将新的参量引入 SSB 模型已经成为 SSB 经验模型研究的一个新方向，本文提出的 ILLR 估计器可以在不影响 SSB 估计结果条件下大幅度降低估计 SSB 所需要的时间，这为实现高维非参数 SSB 模型的实时运算奠定了基础。

参 考 文 献

[1] 王磊. 高精度卫星雷达高度计数据处理技术研究[D]. [博士论文], 中国科学院(国家空间科学中心), 2015.

WANG L. Study on the data processing for high precision satellite radar altimeter[D]. [Ph.D. dissertation], National Space Science Center, Chinese Academy of Sciences, 2015.

[2] 王磊, 许可, 史灵卫, 等. 一种消除合成孔径雷达高度计延迟校正中残余误差的新算法及仿真验证[J]. 电子与信息学报, 2015, 37(11): 2713-2718. doi: 10.11999/JEIT150282.

WANG L, XU K, SHI L W, *et al.* A new range migration correction algorithm and its simulation for SAR altimeter[J]. *Jounal of Electronics & Information Technology*, 2015, 37(11): 2713-2718. doi: 10.11999/JEIT150282.

[3] 刘亚龙. HY-2 雷达高度计海面高度定标技术研究[D]. [博士论文] 中国海洋大学, 2014.

LIU Y L. Calibration technology for HY-2 radar altimeter sea surface height[D]. [Ph.D. dissertation], Ocean University of China, 2014.

[4] GASPAR P, OGOR F, LETRAON P Y, *et al.* Estimating the sea-state bias of the topex and poseidon altimeters from crossover differences[J]. *Journal of Geophysical Research-Oceans*, 1994, 99(C12): 24981-24994. doi: 10.1029/94JC01430.

[5] PRANDI P, PHILIPPS S, PIGNOT V, *et al.* SARAL/AltiKa global statistical assessment and cross-calibration with Jason-2[J]. *Marine Geodesy*, 2015, 38(10): 297-312. doi: 10.1080/01490419.2014.995840.

[6] TRAN N, VANDEMARK D, CHAPRON B, *et al.* New models for satellite altimeter sea state bias correction developed using global wave model data[J]. *Journal of Geophysical Research-Oceans*, 2006, 111(C9): 141-152. doi: 10.1029/2005JC003406.

[7] BONNEFOND P, EXERTIER P, LAURAIN O, *et al.* SARAL/AltiKa absolute calibration from the multi-mission corsica facilities[J]. *Marine Geodesy*, 2015, 38(10): 171-192. doi: 10.1080/01490419.2015.1029656.

[8] GASPAR P and FLOREANS J P. Estimation of the sea state bias in radar altimeter measurements of sea level: Results from a new nonparametric method[J]. *Journal of Geophysical Research-Oceans*, 1998, 103(C08): 15803-15814. doi: 10.1029/98JC01194

[9] CHELTON D B. The sea-state bias in altimeter estimates of sea-level from collinear analysis of topex Data[J]. *Journal of Geophysical Research-Oceans*, 1994, 99(C12): 24995-25008. doi: 10.1029/94JC02113

[10] GASPAR P, LABROUE S, OGOR F, *et al.* Improving nonparametric estimates of the sea state bias in radar altimeter measurements of sea level [J]. *Journal of Atmospheric and Oceanic Technology*, 2002, 19(10): 1690-1707. doi: 10.1175/1520-0426(2002)019<1690:INEOTS>2.0.CO;2.

[11] WANG J X and XIAO Q X. Local polynomial estimation of time-dependent diffusion parameter for discretely observed SDE models[J]. *Filomat*, 2014, 28(4): 871-878. doi: 10.2298/FIL1404871W.

[12] SU L and ULLAH A. Local polynomial estimation of nonparametric simultaneous equations models[J]. *Journal of Econometrics*, 2008, 144(1): 193-218. doi:10.1016/j.jeconom.2008.01.002.

[13] DETTMERING D, SCHWATKE C, and BOSCH W. Global calibration of SARAL/AltiKa using multi-mission sea surface height crossovers[J]. *Marine Geodesy*, 2015, 38(10): 206-218. doi: 10.1080/01490419.2014.988832.

[14] PAIGE C C and SAUNDERS M A. LSQR: An algorithm for sparse linear-equations and sparse least-squares[J]. *Acm Transactions on*

Mathematical Software, 1982, 8(1): 43-71. doi: 10.1145/355984.355989.
[15] FONG D C L and SAUNDERS M. LSMR: An iterative algorithm for sparse least-squares problems[J]. *Siam Journal on Scientific Computing*, 2010, 33(5):2950-2971. doi: 10.1137/10079687X.
[16] TRAN N, VANDEMARK D, LABROUE S, *et al.* Sea state bias in altimeter sea level estimates determined by combining wave model and satellite data[J]. *Journal of Geophysical Research-Oceans*, 2010, 115(C03020):1-7. doi: 10.1029/2009JC005534.
[17] FENG H, YAO S, LI L Y, *et al.* Spline-based nonparametric estimation of the altimeter sea-state bias correction[J]. *IEEE Geoscience and Remote Sensing Letters*, 2010, 7(3): 577-581. doi: 10.1109/LGRS.2010.2041894.

蒋茂飞：男，1989 年生，博士生，研究方向为雷达高度计数据处理.
许 可：男，1963 年生，博士，研究员，博士生导师，主要研究方向为星载雷达高度计系统技术、合成孔径雷达高度计系统技术和信号处理技术.
刘亚龙：男，1985 年生，博士，研究方向为雷达高度计数据处理.
王 磊：男，1986 年生，博士，研究方向为雷达高度计信号处理